AD NUMBER

AD829221

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; Feb 1968. Other requests shall be referred to Dean of Engineering, Air Force Institute of Technology [AFIT-SE], Wright-Patterson AFB, OH 45433.

AUTHORITY

AFIT memo dtd 22 Jul 1971

AIR FORCE INSTITUTE OF TECHNOLOGY

AIR UNIVERSITY
UNITED STATES AIR FORCE

SCHOOL OF ENGINEERING

WRIGHT-PATTERSON AIR FORCE BASE, OHIO

THE EFFECT OF ORDERING ON THE AVERAGE RATE OF CONVERGENCE OF THE POINT SUCCESSIVE OVERRELAXATION ITERATIVE METHOD

THESIS

GSP/PH/68-5

Richard E. Durrett
Capt USAF

AFLC-WPAFB-APR 68 35

THE EFFECT OF ORDERING ON THE AVERAGE RATE OF CONVERGENCE OF THE POINT SUCCESSIVE OVERRELAXATION ITERATIVE METHOD

THESIS

Presented to the Faculty of the School of Engineering of
the Air Force Institute of Technology
Air University
in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

by

Richard E. Durrett, B.S.
Capt USAF

Graduate Space Physics

February 1968

Preface

This study began with an attempt to explain, theoretically, or experimentally, how the increased rate of convergence was achieved for two methods of iteration previously developed at the Air Force Institute of Technology. Analysis of the simpler method applied to point successive overrelaxation was attempted first. Difficulties in applying the theory to the method and achieving the reported increase in convergence rates in actual test runs, were immediately encountered. The length of time needed to finally resolve these difficulties precluded analysis of the other method, which was applied to block successive overrelaxation.

Approximately five hours of computer time on an IBM 7094 computer were used in testing and comparing various scanning techniques with the method previously developed. The method of point successive overrelaxation was applied to the five point finite difference approximations to Laplace's equation and the transient heat transfer equation.

Two scanning methods, which I have elected to call the standard scan and the even-odd scan, were programed along with the method previously developed. It was found that the standard and even-odd scans were iteratively faster as the number of required iterations was increased. Both of these scans are "consistent" and therefore qualify for the maximum asymptotic rate of convergence amongst the group of all possible scans. I have attempted in chapter II to explain explicitly how one determines when a given scan is consistent. The method is due to Varga and can be easily applied once it is understood.

I wish to express my gratitude to Dr. Bernard Kaplan of the Physics Department at the Air Force Institute of Technology for his guidance during this study, and to my wife for typing the thesis.

Richard E. Durrett

Contents

List of Figures and Tables

List of Symbols

$\underline{A}$	matrix of difference equations
$\underline{B}$	Jacobi matrix $\equiv -\underline{D}^{-1}\,\underline{A} + \underline{I}$
$\underline{D}$	strictly diagonal matrix
$\underline{E}$	strictly lower triangular matrix
$\bar{e}$	eigenvector of $\underline{L}_\omega$
$\underline{F}$	strictly upper triangular matrix
h	mesh spacing in x direction
i	mesh index in x direction
j	mesh index in y direction
k	mesh spacing in y direction
$\underline{L}_\omega$	iteration matrix of the S.O.R. iterative method
m	iteration number
n	number of internal mesh points
N	number of mesh spacings
$\underline{P}$	permutation matrix
p	index of a cyclic matrix
r	$\equiv \alpha\,\Delta t/h^2$
α	thermal diffusivity
β	function of heat generation rate, density of material, and heat capacity
γ	difference magnitude $\equiv \max \left\lvert u_{i,j}^{(m+1)} - u_{i,j}^{(m)} \right\rvert$, $1 \leq i, j \leq N - 1$
$\bar{\epsilon}$	error vector
λ	eigenvalue of $\underline{L}_\omega$
μ	eigenvalue of the Jacobi matrix
$\rho(\underline{A})$	spectral radius of matrix $\underline{A}$
ω	relaxation factor
ω_b	optimum relaxation factor

Abstract

This study investigates the effect of ordering of finite difference equations on the average rate of convergence of the point successive overrelaxation iterative method. The method is applied to the two dimensional five point implicit finite difference representation of Laplace's equation and the diffusion equation of transient heat transfer.

The average rate of convergence for a specified number of iterations is found to depend on the magnitude of the projection of the initial error vector onto the dominant eigenvector of the point successive overrelaxation iteration matrix. Although the asymptotic rate of convergence has been proven to be the same, and optimum, for all consistent orderings, the average rate of convergence for a specific number of iterations is shown to be different for two consistent orderings. For a reduction in the initial error vector by three orders of magnitude, a difference of 20% in average convergence rates is achieved. An increase in the orders of magnitude reduction produces a proportionate decrease in the per cent difference between the two consistent scans.

I. Introduction

With the growth in speed and complexity of modern high speed digital computers has come an increase in interest in finding or approximating the solution to partial differental equations in several variables by use of these machines. It is naturally desirable to find a method which will yield a solution to the matrix problems arising from discrete approximations to partial differential equations in the mimimum time on the computer. This will allow more and/or larger problems to be solved with the required accuracy.

From the early work of Young (Ref 7) and Frankel (Ref 4) in 1950, the method of point successive overrelaxation (to use Young's nomenclature) has been expanded to cover a wide range of matrix equations through the concept of p-cyclic matrices treated by Varga (Ref 6). This theory also applies to the newer method of block successive overrelaxation. The mathematical theory, although it does provide a good general framework and basis for direction in research, is not sufficiently complete to determine the optimum method.

For large problems the most successful technique to date employs the use of an implicit approximation to the differential equation wherein the corresponding difference equations are solved repeatedly until two successive iterations yield solutions which differ only by a small predetermined amount. The discrete values so obtained are then taken as the solution of the differential equation at the respective points. Various methods have been developed which use this general approach.

Previous work at the Air Force Institute of Technology by Cudahy (1965) (Ref 2) with point successive overrelaxation and Wright (1967) (Ref 10) with block successive overrelaxation showed, by an essentially empirical approach, that the ordering of mesh points, or blocks of mesh points, for solution of the difference equations can have a significant effect on the rate of convergence of the iterative method.

Purpose

The objective of the present study was to investigate the effect of ordering, or method of mesh point scanning, on the average and asymptotic rates of convergence of the point successive overrelaxation iterative method, and to account for the differences in experimentally observed rates of convergence for different scanning methods.

Method

The point successive overrelaxation iterative method was applied to the five point finite difference approximations to Laplace's equation and the heat equation. The equations were applied over a square region in two dimensions so that the optimum relaxation factor could be predicted theoretically for Laplace's equation, and so that the present data could be correlated with the previous work of Cudahy (Ref 2) for the heat equation. The ability to predict the optimum relaxation factor greatly reduces the computer time necessary to obtain the desired data. For the heat equation solutions were obtained with several relaxation factors for each time step. The optimum relaxation factor is different for each time step.

The even-odd scan (denoted odd-even parity by Forsythe and Wasow) (Ref 3:200) and the method developed by Cudahy (Ref 2:34) were compared

with the standard scan used by Young (Ref 8). The results of those comparisons are presented in chapter III. Chapter IV contains a discussion of the results of this study in terms of the theory presented in chapter II.

II. Theory

Point Successive Overrelaxation

To derive the point successive overrelaxation iterative method consider a system of linear equations

$$\sum_{j=1}^{n} a_{i,j} x_j = k_i, \qquad 1 \leq i \leq n \tag{1}$$

where the $a_{i,j}$ are elements of the n x n complex matrix $\underline{A}$. This can be written in matrix notation as

$$\underline{A}\,\bar{x} = \bar{k} \tag{2}$$

where $\bar{k}$ is a given column vector. The solution to this equation exists and is unique if and only if the matrix $\underline{A}$ is nonsingular (Ref 6:56). The solution vector can then be written explicitly as

$$\underline{A}^{-1}\,\bar{k} \tag{3}$$

The diagonal elements $a_{i,i}$ of $\underline{A}$ are now assumed to be nonzero complex numbers.

The matrix $\underline{A}$ can be expressed as

$$\underline{A} = \underline{D} - \underline{E} - \underline{F} \tag{4}$$

where $\underline{D}$ is a strictly diagonal matrix, $\underline{E}$ is a strictly lower triangular matrix, and $\underline{F}$ is a strictly upper triangular matrix. With this splitting of the matrix $\underline{A}$, equation (2) can be written

$$(\underline{D} - \underline{E})\,\bar{x} = \underline{F}\,\bar{x} + \bar{k} \tag{5}$$

Multiplying both sides of the equation by a quantity ω, called the relaxation factor, gives

$$(\omega \underline{D} - \omega \underline{E})\, \bar{x} = \omega \underline{F}\, \bar{x} + \omega \bar{k} \tag{6}$$

Adding the matrix product $\underline{D}\, \bar{x}$ to both sides and rearranging terms gives

$$(\underline{D} - \omega \underline{E})\, \bar{x} = \left[(1 - \omega)\, \underline{D} + \omega \underline{F}\right] \bar{x} + \omega \bar{k} \tag{7}$$

This equation suggests the following iterative method:

$$(\underline{D} - \omega \underline{E})\, \bar{x}^{(m+1)} = \left[(1 - \omega)\, \underline{D} + \omega \underline{F}\right] \bar{x}^{(m)} + \omega \bar{k}, \quad m \geq 0 \tag{8}$$

where the components of $\bar{x}^{(0)}$ are the initial estimates of the unique solution of $\bar{x}$ of equation (2). As $\underline{D} - \omega \underline{E}$ is nonsingular for any choice of ω, and with the definitions $\underline{L} \equiv \underline{D}^{-1}\underline{E}$ and $\underline{U} \equiv \underline{D}^{-1}\underline{F}$, equation (8) can be written

$$\begin{aligned} \bar{x}^{(m+1)} = {} & (\underline{I} - \omega \underline{L})^{-1} \left[(1 - \omega)\underline{I} + \omega \underline{U}\right] \bar{x}^{(m)} \\ & + \omega\, (\underline{I} - \omega \underline{L})^{-1}\, \underline{D}^{-1}\, \bar{k} \end{aligned} \tag{9}$$

This iterative method is called the point successive overrelaxation iterative method and will be denoted the S.O.R. iterative method. The matrix

$$\underline{\mathcal{L}}_{\omega} \equiv (\underline{I} - \omega \underline{L})^{-1} \left[(1 - \omega)\, \underline{I} + \omega \underline{U}\right] \tag{10}$$

is called the iteration matrix of the S.O.R. iterative method.

Spectral Radius. The magnitude of the largest of the eigenvalues of the iteration matrix $\underline{\mathcal{L}}_{\omega}$ is termed the spectral radius. The eigenvector associated with this eigenvalue will be called the dominant eigenvector of the iteration matrix. Varga (Ref 6:13) has shown that

the S.O.R. iterative method is convergent if and only if the spectral radius

$$\rho(\underline{L}_\omega) < 1 \tag{11}$$

This condition is assured provided (a) that the matrix $(\underline{D} - \omega \underline{E})$ is nonsingular and its inverse is non-negative, and (b) that the matrix $\left[(1-\omega)\underline{D} + \omega \underline{F}\right]$ is non-negative (Ref 6:89).

Error Vectors. The error vectors $\bar{\epsilon}^{(m)}$ associated with the S.O.R. iterative method are defined by

$$\bar{\epsilon}^{(m)} = \bar{x}^{(m)} - \bar{x}, \qquad m \geq 0 \tag{12}$$

where $\bar{x}$ is the unique vector solution of equation (2). By equation (9) the error vectors can be expressed as

$$\bar{\epsilon}^{(m)} = \underline{L}_\omega \bar{\epsilon}^{(m-1)} = \ldots = \underline{L}_\omega^m \bar{\epsilon}^{(0)} \tag{13}$$

Assuming that the eigenvectors of $\underline{L}_\omega$ form a complete set, the initial error vector $\bar{\epsilon}^{(0)}$ can be expressed in terms of them as

$$\bar{\epsilon}^{(0)} = a_1\bar{e}_1 + a_2\bar{e}_2 + \ldots + a_n\bar{e}_n \tag{14}$$

where the eigenvectors $\bar{e}_1, \bar{e}_2, \ldots, \bar{e}_n$ are assumed to be normalized. The $a_1, a_2, \ldots, a_n$ represent the magnitudes of the components of the initial error vector. Multiplying equation (14) by $\underline{L}_\omega$ yields

$$\bar{\epsilon}^{(1)} = \underline{L}_\omega \bar{\epsilon}^{(0)} = a_1 \underline{L}_\omega \bar{e}_1 + a_2 \underline{L}_\omega \bar{e}_2 + \ldots a_n \underline{L}_\omega \bar{e}_n \tag{15}$$

which by the eigenvalue equation $\underline{L}_\omega \bar{e}_i = \lambda_i \bar{e}_i$ reduces to

$$\bar{\epsilon}^{(1)} = a_1 \lambda_1 \bar{e}_1 + a_2 \lambda_2 \bar{e}_2 + \ldots + a_n \lambda_n \bar{e}_n \tag{16}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of $\underline{L}_\omega$. Continuing in this fashion the error vector at each iteration is given by

$$\begin{aligned} \bar{\epsilon}^{(2)} &= a_1 \lambda_1^2 \bar{e}_1 + a_2 \lambda_2^2 \bar{e}_2 + \ldots + a_n \lambda_n^2 \bar{e}_n \\ &\vdots \qquad\qquad \vdots \\ \bar{\epsilon}^{(m)} &= a_1 \lambda_1^m \bar{e}_1 + a_2 \lambda_2^m \bar{e}_2 + \ldots + a_n \lambda_n^m \bar{e}_n \end{aligned} \tag{17}$$

Assuming that the eigenvalues are ordered by magnitude such that

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| \tag{18}$$

then

$$|\lambda_1^m| > |\lambda_2^m| > \cdots > |\lambda_n^m| \tag{19}$$

For sufficiently large m, $|\lambda_1^m| >> |\lambda_2^m|$ and

$$\bar{\epsilon}^{(m)} \cong a_1 \lambda_1^m \bar{e}_1 \qquad \text{(Ref 1:115)} \tag{20}$$

By definition the spectral radius $\rho(\underline{L}_\omega)$ is equal to the magnitude of the largest eigenvalue λ_1 of $\underline{L}_\omega$, thus it can be seen that the error vector will vanish as $m \longrightarrow \infty$ if and only if $\rho(\underline{L}_\omega) < 1$. For sufficiently large m the magnitude of the error vector at each iteration is primarily a function of the spectral radius and the quantity a_1 defined by equation (14). Both of these quantities are determined by the choice of the iteration matrix $\underline{L}_\omega$, which in turn is determined by the choice of the relaxation factor ω and the order in which the equations (i) are solved. A particular ordering has been implied in this development by the splitting of matrix $\underline{A}$ in equation (4). This ordering is termed the natural or standard ordering.

Rate of Convergence. There is some disagreement in the literature concerning the definition of rate of convergence. Young, in 1950, defined the rate of convergence as $- \ln \rho (\underline{L}_\omega)$ (Ref 7:94). Varga, however, has chosen to call this the asymptotic rate of convergence since it is realized only as $m \longrightarrow \infty$. Varga then defines the average rate of convergence as

$$R(\underline{A}^m) = - \ln \left[\left(\left|\left| \underline{A}^m \right|\right| \right)^{1/m} \right] \tag{21}$$

for m iterations, where $\left|\left| \underline{A} \right|\right|$ denotes the spectral norm of $\underline{A}$ (Ref 6:62). If two iteration matrices, $\underline{A}$ and $\underline{B}$, have different average rates of convergence for m iterations such that $R(\underline{A}^m) < R(\underline{B}^m)$ then matrix $\underline{A}$ is said to be iteratively faster for m iterations (Ref 6:62). Varga's definitions will be used throughout this paper. The average rate of convergence and the average error reduction rate are equivalent.

Consistent Ordering. Varga (Ref 6:125) has shown that any ordering of equations (1) which is "consistent" with the natural ordering gives rise to an iteration matrix $\underline{L}_\omega$ whose spectral radius is equal to that for $\underline{L}_\omega$ given in equation (10) for the same ω. Furthermore, Varga has shown that this spectral radius is smaller than the spectral radius for any other ordering which is not consistent with the natural ordering. Hence, for sufficiently large m the rate of error vector reduction will be greatest for consistent orderings.

Let the n x n point Jacobi matrix $\underline{B}$ be defined by

$$\underline{B} \equiv -\underline{D}^{-1} \underline{A} + \underline{I} \tag{22}$$

where matrix $\underline{D}$ is defined in equation (4), and matrix $\underline{A}$ is defined in equation (2). Then the matrix $\underline{A}$ is consistently ordered if all the

eigenvalues of the matrix

$$\underline{B}(\eta) \equiv \eta \underline{L} + \eta^{-(p-1)} \underline{U} \tag{23}$$

derived from the matrix $\underline{B} = \underline{L} + \underline{U}$, where $\underline{L}$ and $\underline{U}$ are respectively strictly upper and strictly lower triangular matrices, are independent of η, for $\eta \neq 0$, provided that $\underline{A}$ is a p-cyclic matrix (Ref 6:101). The matrix $\underline{A}$ is defined as p-cyclic if the matrix $\underline{B}$ of (22) is weakly cyclic of index p (≥ 2) (Ref 6:99). The matrix $\underline{B}$ is weakly cyclic of index p (Ref 6:39) if there exists an n x n permutation matrix $\underline{P}$ such that $\underline{PBP}^T$ is of the form

$$\underline{PBP}^T = \begin{bmatrix} 0 & 0 & \cdots & 0 & b_{1,p} \\ b_{2,1} & 0 & & 0 & 0 \\ 0 & b_{3,2} & & 0 & 0 \\ \cdot & & \ddots & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ 0 & 0 & \cdots & b_{p,p-1} & 0 \end{bmatrix} \tag{24}$$

The eigenvalues of the point successive overrelaxation iteration matrix $\mathcal{L}_\omega$ derived from a consistently ordered p-cyclic matrix $\underline{A}$ are related to the eigenvalues of the associated Jacobi matrix $\underline{B}$ by

$$(\lambda + \omega - 1)^p = \lambda^{p-1} \omega^p \mu^p \tag{25}$$

where λ is a nonzero eigenvalue of $\mathcal{L}_\omega$, ω is the relaxation factor, and μ is an eigenvalue of $\underline{B}$ (Ref 6:106). That such a relationship exists is itself interesting, but its importance lies in the fact that knowledge of the eigenvalues of the matrix $\underline{B}$ allows determination of the eigenvalues of $\mathcal{L}_\omega$ for a consistent ordering.

Directed Graph. Fortunately, it is not necessary to apply the definition directly to determine whether an ordering is consistent.

It can be shown (Ref 6:121) that an ordering is consistent if it leads to a matrix $\underline{B}$ whose directed graph of type 2 has an equal number of major and minor paths on every closed path. A directed graph of type 2 for the matrix $\underline{B} \equiv (b_{i,j})$ is constructed so that if $b_{i,j} \neq 0$ then a path from node P_i to the node P_j is drawn and denoted by a double-arrow (see Fig. 1) only if $j > i$; otherwise, a single-arrowed path is drawn. The former paths are called major paths; the other paths are called minor paths. It is also possible to use this graph to determine when the matrix $\underline{A}$ is p-cyclic; or equivalently, when the point Jacobi matrix $\underline{B}$ is weakly cyclic of index p (Ref 6:100). If the greatest common divisor of the lengths of closed paths of the graph is p then the matrix $\underline{B}$ is weakly cyclic of index p, where the length of a path is defined as the number of nodes reached in traversing the path.

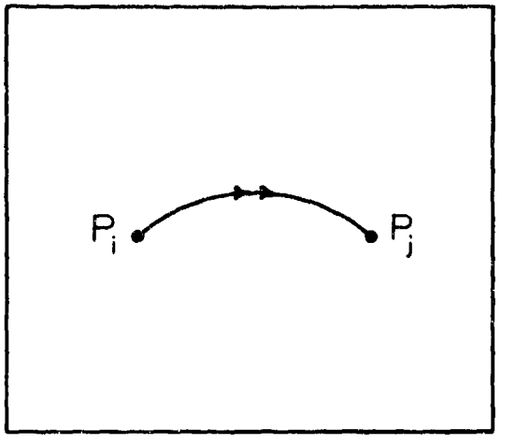

Fig. 1 Type 2 Directed Graph

Optimum Relaxation Factor. As stated previously the spectral radius of the iteration matrix $\underline{L}_\omega$ depends on the choice of the relaxation factor ω, as well as the ordering of equations (1). Ostrowski has shown that the S.O.R. iterative method is convergent for all ω such that $0 < \omega < 2$ (Ref 6:77). The optimum ω, in the sense that $\rho(\underline{L}_\omega)$ is minimized, can be related to the largest of the eigenvalues of the matrix $\underline{B}$ of equation (22), for the special case $p = 2$, as

$$\omega_b = \frac{2}{1 + \sqrt{1 - \rho^2(\underline{B})}} \tag{26}$$

where ω_b denotes the optimum ω (Ref 6:110).

Thus, for 2-cyclic matrices, the problem of finding the ordering and relaxation factor which yields the greatest rate of error vector reduction in the limit as $m \longrightarrow \infty$ has been reduced theoretically to choosing any consistent ordering and finding the largest eigenvalue of the associated point Jacobi matrix $\underline{B}$. The theory still allows for the possibility, however, that the average rate of convergence may be different for different consistent orderings, and hence, that the approach to asymptotic convergence may differ between consistent orderings. Also, the projection of the initial error vector onto the dominant eigenvector of the iteration matrix, expressed as the quantity a_1 in equation (20), may be different for different orderings even if all are consistent. The extent to which this may be of practical importance has been examined in this study by applying the point successive overrelaxation iterative method to the finite difference approximations to elliptic and parabolic partial differential equations.

Difference Equations

To derive these difference equations consider the general second order, linear, partial differential equation

$$A \frac{\partial^2 u}{\partial x^2} + B \frac{\partial^2 u}{\partial x \partial y} + C \frac{\partial^2 u}{\partial y^2} + D \frac{\partial u}{\partial x} + E \frac{\partial u}{\partial y} + Fu = G \qquad (27)$$

where A, B, ... , G are constants or function of x and y only. This equation is classified as elliptic, parabolic, or hyperbolic in a domain of the xy plane as the values of the function $B^2 - 4\,A\,C$ are negative, zero, or positive, respectfully, throughout the domain. Equations of the hyperbolic type were not considered in this study.

Laplace's Equation. Laplace's equation is an example of the elliptic type, and this equation was chosen for study because the eigenvalues of the iteration matrix of the associated difference equations can be found theoretically. This allows precise determination of the optimum overrelaxation factor prior to actual solution of the difference equations.

To derive a finite difference approximation to Laplace's equation

$$\frac{\partial^2 u(x,y)}{\partial x^2} + \frac{\partial^2 u(x,y)}{\partial y^2} = 0 \tag{28}$$

subdivide the x-y plane into sets of equal rectangles of sides $\delta x = h$, $\delta y = k$, as shown in Fig. 2, and let the coordinates (x,y) of a representative mesh point p be

$$x = ih, \qquad y = jk \tag{29}$$

where i and j are integers (Ref 5:7).

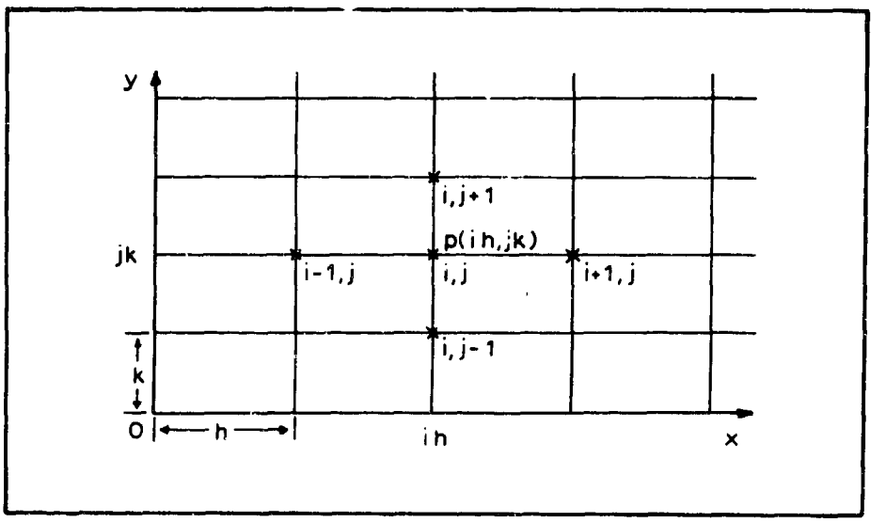

Fig. 2 Sample Mesh

Denote the value of u at p by

$$u_p = u(\quad jk) = u_{i,j} \tag{30}$$

Expand the function u(x,y) in the x direction about the point x_p in a Taylor's expansion as

$$u(x_p \pm h, y_p) = u(x_p, y_p) \pm h \left.\frac{\partial u}{\partial x}\right|_{x_p,y_p} + \tfrac{1}{2} h^2 \left.\frac{\partial^2 u}{\partial x^2}\right|_{x_p,y_p} \pm \frac{1}{6} h^3 \left.\frac{\partial^3 u}{\partial x^3}\right|_{x_p,y_p} + \ldots \tag{31}$$

Addition of the expansions for $u(x_p + h, y_p)$ and $u(x_p - h, y_p)$ gives

$$u(x_p + h, y_p) + u(x_p - h, y_p) = 2u(x_p, y_p) + h^2 \left.\frac{\partial^2 u}{\partial x^2}\right|_{x_p,y_p} + O(h^4) \tag{32}$$

where $O(h^4)$ denotes terms containing fourth and higher powers of h. Assuming that these terms are negligible compared with the lower powers of h, it follows that

$$\left.\frac{\partial^2 u}{\partial x^2}\right|_{x_p,y_p} = \frac{u(x_p + h, y_p) - 2u(x_p, y_p) + u(x_p - h, y_p)}{h^2} \tag{33}$$

In terms of the notation introduced in equation (29) this can be expressed as

$$\left.\frac{\partial^2 u}{\partial x^2}\right|_{p} = \frac{u\left[(i+1)h, jk\right] - 2u\left[ih, jk\right] + u\left[(i-1)h, jk\right]}{h^2} \tag{34}$$

or

$$\left.\frac{\partial^2 u}{\partial x^2}\right|_{i,j} = \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2} \tag{35}$$

The derivation for $\partial^2 u / \partial y^2$ is similar and gives

$$\left.\frac{\partial^2 u}{\partial y^2}\right|_{i,j} = \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{k^2} \tag{36}$$

Both equations (35) and (36) are approximations to the derivatives, and are called finite difference approximations. The magnitude of the error introduced by these equations is reduced as the mesh size is decreased.

Substituting the expressions for the derivatives, equations (35) and (36), gives the following approximation to equation (28), Laplace's equation,

$$\frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2} + \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{k^2} = 0 \tag{37}$$

Assuming for simplicity of analysis that the mesh spacings are taken equal in the x and y directions, that is, that h = k, equation (37) reduces to

$$4u_{i,j} - (u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}) = 0 \tag{38}$$

for each mesh point. This expression is called a five-point formula since each mesh point (x_i, y_j) is coupled to at most four other adjacent mesh points. This system of linear equations can be written

$$\underline{A}\,\bar{u} = \bar{k} \tag{39}$$

where the column vector $\bar{k}$ contains the values of the function u specified on the boundaries of a rectangular region in the xy plane.

The derivation leading to equation (39) shows that $\underline{A}$ is a real symmetric matrix with positive diagonal entries and non-positive off-diagonal entries. Moreover, it can be shown that $\underline{A}$ is irreducibly diagonally dominant, so that $\underline{A}$ is positive definite (Ref 6:187). This implies that $\underline{A}^{-1} > \underline{0}$, where $\underline{0}$ is the null matrix (Ref 6:85).

The associated point Jacobi matrix $\underline{B}$, defined by equation (22), can be shown to be a non-negative, irreducible cyclic matrix of index 2 with real eigenvalues, and with $\rho(\underline{B}) < 1$ (Ref 6:188).

Since all of the diagonal entries of the matrix $\underline{A}$ are 4, the point Jacobi matrix $\underline{B}$ associated with $\underline{A}$ can be expressed simply as

$$\underline{B} = \underline{I} - \tfrac{1}{4}\underline{A} \tag{40}$$

All the eigenvalues and eigenvectors of the matrix $\underline{B}$ can be obtained explicitly for the finite difference approximation to Laplace's equation in a rectangle. Since $\underline{B}$ is non-negative and irreducible, it has a unique positive eigenvector with eigenvalue $\rho(\underline{B})$, which for a unit square is given by

$$\rho(\underline{B}) = \cos(h\pi) \qquad \text{(Ref 6:203)} \tag{41}$$

For the point successive overrelaxation iterative method the optimum ω can thus be written, by equation (26), as

$$\omega_b = \frac{2}{1 + \sqrt{1 - \rho^2(\underline{B})}} = \frac{2}{1 + \sin(h\pi)} \tag{42}$$

for all consistent orderings.

Heat Equation. The derivation of the five point finite difference approximation to the diffusion equation of transient heat transfer

is similar to that for Laplace's equation. The mesh spacings and notation defined in equation (30) and illustrated in Fig. 2 are again adopted.

In two dimensions the heat equation for an isotropic, homogeneous material can be written as

$$\frac{\partial U}{\partial t} = \alpha \left[\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} \right] + \beta \tag{43}$$

where

$U(x,y,t)$ is the temperature,

α is the thermal diffusivity, and

β depends on the rate of heat generation, the density of the material, and the heat capacity.

It can be seen that the steady state heat equation with no heat generation reduces to Laplace's equation, while the transient heat equation is parabolic.

With the assumption of uniform mesh spacing, equal to h, the second partials of U with respect to x and y at time $t + \Delta t$ can be replaced with finite differences similar to equations (35) and (36) while the first partial of U with respect to t can be found from the Taylor's expansion

$$U(x_p,y_p,t) = U(x_p,y_p,t + \Delta t) - \Delta t \left. \frac{\partial U}{\partial t} \right|_{x_p,y_p,t + \Delta t} + \tfrac{1}{2}(\Delta t)^2 \left. \frac{\partial^2 U}{\partial t^2} \right|_{x_p,y_p,t + \Delta t} - \ldots \tag{44}$$

Assuming that higher powers of Δt can be neglected, this becomes

$$\left.\frac{\partial U}{\partial t}\right|_{x_p,y_p,t+\Delta t} = \frac{U(x_p,y_p,t+\Delta t) - U(x_p,y_p,t)}{\Delta t} \tag{45}$$

Using these finite difference equations at time $t + \Delta t$, equation (43) can be expressed as

$$\frac{U(i,j,t+\Delta t) - U(i,j,t)}{\Delta t} = \alpha \frac{U(i+1,j,t+\Delta t) + U(i-1,j,t+\Delta t)}{h^2} + \alpha \frac{U(i,j+1,t+\Delta t) + U(i,j-1,t+\Delta t)}{h^2} - \frac{4\alpha}{h^2} U(i,j,t+\Delta t) + \beta \tag{46}$$

Introducing the notation $r \equiv \alpha \Delta t/h^2$ this equation becomes

$$(1 + 4r)\, U(i,j,t+\Delta t) = r\left[U(i+1,j,t+\Delta t) + U(i-1,j,t+\Delta t) + U(i,j+1,t+\Delta t) + U(i,j-1,t+\Delta t)\right] + U(i,j,t) + \beta \Delta t \tag{47}$$

or

$$U(i,j,t+\Delta t) = \frac{r}{1+4r}\left[U(i+1,j,t+\Delta t) + U(i-1,j,t+\Delta t) + U(i,j+1,t+\Delta t) + U(i,j-1,t+\Delta t)\right] + \frac{1}{1+4r}\left[U(i,j,t) + \beta \Delta t\right] \tag{48}$$

This system of linear equations can be written

$$\underline{A}\, \overline{U} = \overline{K} \tag{49}$$

where $\overline{K}$ includes the given boundary and initial conditions as well as the heat generation rate. Again, as for Laplace's equation, $\underline{A}$ is irreducibly diagonally dominant, positive definite, and thus $\underline{A}^{-1} > \underline{0}$.

The associated point Jacobi matrix $\underline{B}$ is a non-negative, irreducible cyclic matrix of index 2 with real eigenvalues, and with $\rho(\underline{B}) < 1$.

The eigenvalues of the matrix $\underline{B}$, however, cannot be determined explicitly for this equation. Thus the spectral radius $\rho(\underline{B})$ must be estimated, or the optimum ω determined experimentally, for a given problem.

Ordering

The standard method of ordering of the difference equations, such as equation (38), for solution by an iterative method can be illustrated for a small sample mesh, as shown in Fig. 3. The node (i = 1, j = 1) is assigned the first position in the ordering, the node (1,2) the second, and (1,3) the third. The "scan" is then continued in a similar manner for the row i = 2 giving node (2,1) the fourth position in the ordering, node (2,2) the fifth and so on throughout the mesh. If each node within the mesh is assigned a number as shown in Fig. 3, the standard ordering may be expressed as $\left[1,2,3,4,5,6,7,8,9\right]$, where the first position within the brackets designates which point is to be solved first, the second position designates which is to be solved second and so on.

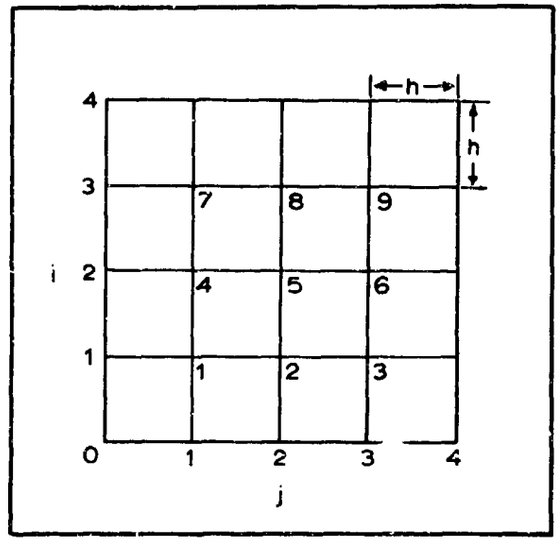

Fig. 3 Mesh Numbering

With this notation a different ordering, called the even-odd scan, can be expressed $\left[2,4,6,8,1,3,5,7,9\right]$. Again, the numbers within the brackets denote the points as labled in Fig. 3, while the positions within the brackets denote the order in which the points are taken for solution, with the first position first, the second, second, and so on. When the point occupying the last position is solved for, the iteration is complete. The ordering is also called the "scan".

The definition of a consistent ordering can now be illustrated for the two scans defined above. Take, for example, the finite difference approximation to Laplace's equation in a square. The mesh and point numberings shown in Fig. 3 will be used. Equation (38) can be written in matrix notation as

$$\underline{A}\,\bar{u} = \bar{k} \tag{50}$$

where the column vector $\bar{k}$ contains the values of the function u specified on the boundaries of the mesh. The matrix $\underline{A}$ is given by

$$\underline{A} = \begin{bmatrix} 1 & -\frac{1}{4} & 0 & -\frac{1}{4} & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{4} & 1 & -\frac{1}{4} & 0 & -\frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{4} & 1 & 0 & 0 & -\frac{1}{4} & 0 & 0 & 0 \\ -\frac{1}{4} & 0 & 0 & 1 & -\frac{1}{4} & 0 & -\frac{1}{4} & 0 & 0 \\ 0 & -\frac{1}{4} & 0 & -\frac{1}{4} & 1 & -\frac{1}{4} & 0 & -\frac{1}{4} & 0 \\ 0 & 0 & -\frac{1}{4} & 0 & -\frac{1}{4} & 1 & 0 & 0 & -\frac{1}{4} \\ 0 & 0 & 0 & -\frac{1}{4} & 0 & 0 & 1 & -\frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{4} & 0 & -\frac{1}{4} & 1 & -\frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{4} & 0 & -\frac{1}{4} & 1 \end{bmatrix} \tag{51}$$

The associated Jacobi matrix $\underline{B}$ is defined by

$$\underline{B} = \underline{I} - D^{-1}\,\underline{A} \tag{52}$$

and is given by

$$\underline{B} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \tag{53}$$

To determine whether an ordering is consistent, the matrix $\underline{B}$ is transformed by the permutation matrix $\underline{P}$ corresponding to the ordering, by a similarity transformation of the form $\underline{PBP}^{-1} = \underline{PBP}^{T}$ (Ref 3:244). The permutation matrix $\underline{P}$ corresponding to the standard ordering $\left[1,2,3,4,5,6,7,8,9\right]$ is simply $\underline{P} = \underline{I}$ and hence the transformation yields the matrix $\underline{B}$. Now, the ordering is consistent if the directed graph of type 2 of the matrix $\underline{PEP}^{T}$ has an equal number of major and minor paths. The directed graph of type 2 for matrix $\underline{B}$ of equation (53) is shown in Fig. 4.

Many useful properties can be deduced from this graph. The graph is "strongly connected", that is, there is a path (with arrows) from each point to every other point. Hence the matrix $\underline{B}$ (and therefore $\underline{A}$)

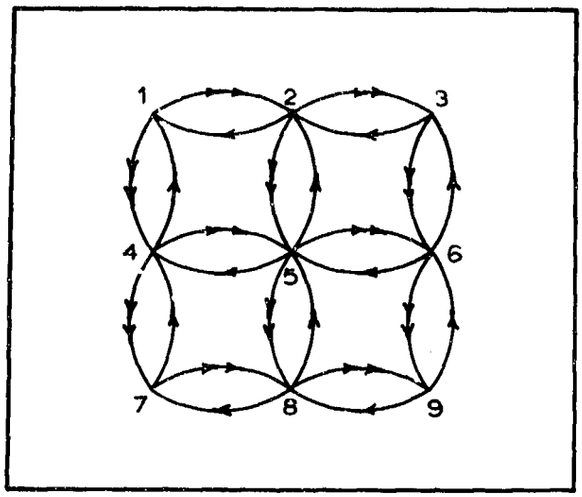

Fig. 4 Directed Graph of Type 2 for the Standard Scan

is irreducible (Ref 6:20). The greatest common divisor of closed paths is 2, hence the matrix $\underline{B}$ is weakly cyclic of index 2 (Ref 6:101). By definition, then, the matrix $\underline{A}$ is 2-cyclic (Ref 6:99). The number of major (2-arrowed) and minor (1-arrowed) paths on every closed path is equal; therefore, the matrix $\underline{B}$ is consistently ordered. Since the same conclusions hold for $\underline{PBP}^T$, the standard scan is consistent.

The permutation matrix $\underline{P}$ for the even-odd scan $\left[2,4,6,8,1,3,5,7,9\right]$ is given by

$$\underline{P} = \begin{bmatrix} 0&1&0&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&0&1&0 \\ 1&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0 \\ 0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0&1 \end{bmatrix} \tag{54}$$

The directed graph of type 2 for $\underline{PBP}^T$, where $\underline{B}$ is given by equation (53), is shown in Fig. 5. The number of major and minor paths on each

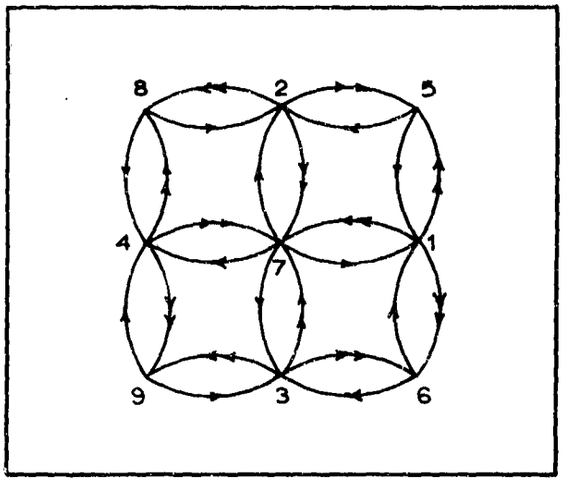

Fig. 5 Directed Graph of Type 2 for the Even-Odd Scan

closed path is equal, and therefore the even-odd scan is consistent.

III. Experimental Investigation and Results

In this investigation the standard scan and the even-odd scan were programmed in Fortran IV for execution on an IBM 7094 computer. The scans were applied to the five point finite difference approximations to Laplace's equation and the heat equation in a square with a uniform mesh spacing in both directions. The method of point successive overrelaxation was used.

These two scans were compared to each other and to a method developed and tested by Cudahy in 1965 (Ref 2). Cudahy applied his method, which was designed to accelerate the convergence of the standard point successive overrelaxation iterative method, to the five point finite difference approximation to the diffusion equation of transient heat transfer, using point successive overrelaxation. He compared his method to the standard scan and found that his method was iteratively faster, the per cent difference depending on various factors, including the time step and heat generation rate. He did not program the even-odd scan.

One of the first objectives of the present study was to explain, from an examination of the theory of point successive overrelaxation, why Cudahy's method was faster than the standard scan. This proved to be somewhat difficult, however, since the theory that has been developed for point successive overrelaxation deals with cyclic, stationary iterations, while Cudahy's method strictly speaking is neither cyclic nor stationary.

An iterative procedure is said to be cyclic when the ordering is repeated, without change, on each iteration. The method developed by

Cudahy employs four different orderings (each of which, considered by itself, is consistent). These four scans are applied repetitively, in the same order, until the process converges to a predetermined limit. The four orderings can be illustrated in terms of the mesh points labeled in Fig. 3. The first ordering is $[1,2,3,7,8,9,4,5,6]$; the second ordering is $[3,2,1,9,8,7,6,5,4]$; the third ordering is $[1,4,7,3,6,9,2,5,8]$; the fourth ordering is $[7,4,1,9,6,3,8,5,2]$. The center row or column is always solved last.

An iterative process is termed stationary if the relaxation factor applied is independent of the iteration number. In his method, Cudahy applied two different relaxation factors on each iteration. Although the two relaxation factors are the same on each subsequent iteration, the use of four different orderings makes the relaxation factor applied to certain points of the mesh a function of the iteration number, and hence the method is non-stationary.

One of the difficulties that this presents can be seen by examining the assumptions used in the error vector analysis leading to equation (20) of chapter II. It was assumed there that the same iteration matrix $\underline{L}_\omega$ of equation (10) was applied on each iteration, that is, that the method was cyclic and stationary. For Cudahy's method there are four different iteration matrices, which do not have the same set of eigenvectors, and thus equation (20) does not apply.

Another difficulty introduced by the use of four iteration matrices is encountered in predicting the asymptotic rate of convergence. The rate is affected by a coupling between the iteration matrices which cannot be exactly known without knowledge of the full range of eigenvalues and eigenvectors of each matrix. Determination

of the eigenvalues and eigenvectors of the iteration matrices, which is a formidable problem in itself, is complicated by the fact that explicit representation of the iteration matrices requires, for a 51 by 51 mesh, the direct inversion of a 2401 by 2401 matrix for each iteration matrix. This type of problem is avoided in the theory of point successive overrelaxation by equation (25) which specifies the eigenvalues of the S.O.R. iteration matrix in terms of the eigenvalues of the associated Jacobi matrix $\underline{B}$, which can be explicitly represented from equation (22).

Since the theory provides no immediate answer for these problems, the computer was used to compare Cudahy's method experimentally with the standard and even-odd scans.

Laplace's Equation

The finite difference approximation to Laplace's equation, for a uniform mesh spacing, given by equation (38), can be solved by the method of point successive overrelaxation derived in chapter II. The point equation corresponding to the matrix equation (9) is given by

$$u(i,j) = (1 - \omega)u(i,j) + \frac{\omega}{4}\left[u(i+1,j) + u(i-1,j) + u(i,j+1) + u(i,j-1)\right] \qquad (55)$$

where the $(m+1)$'s and (m)'s have been suppressed to allow this equation to represent any ordering. The values of the function u within the brackets are to be taken from the present or preceding iteration, depending on the ordering of points for solution.

Equation (55) is applied at each point within the mesh to form a complete iteration. The magnitude of the change in the value of u at

each point is c ted as the iteration progresses, and the largest change is retained. At the end of each iteration this value can be compared to a predetermined number, called the iteration limit. If the largest of the changes in magnitude of the function u, denoted γ, is less than the iteration limit the process may be said to have converged to that iteration limit. This is an easily applied convergence criteria.

In this study the difference magnitude γ was not tested at the end of each iteration to determine convergence, but rather a specified number of iterations were performed with the value of γ printed out for each iteration. If an iteration limit is later specified then the first iteration where γ is less than the specified limit will be the iteration at which the process can be said to have converged.

The standard and even-odd scans where compared to Cudahy's method for a 31 by 31 mesh where the value u = 1000 was specified on the boundaries, and the initial estimate for all points within the mesh was u = 0. The results of that comparison are shown in Table 1 for various iteration limits. It was stated previously that Cudahy's method employs two relaxation factors. The second relaxation factor is applied only to the points in the last row or column of the iteration and is given in terms of the ω listed in Table 1 by

$$\omega' = \frac{\omega(1 + 4r)}{\omega r + 1 + 4r} \qquad \text{(Ref 2:25)} \tag{56}$$

where $r = \alpha \frac{\Delta t}{h^2}$, which gives for Laplace's equation ($\Delta t \longrightarrow \infty$)

$$\omega' = \frac{4\omega}{4 + \omega} \tag{57}$$

Table 1

Number of Iterations Required for Various Iteration Limits for a 31 x 31 Mesh Solving Laplace's Equation

Iteration Limit	Iterations					
	Standard Scan	Even-Odd Scan	Cudahy's Method			
	$\omega = 1.81$	$\omega = 1.81$	$\omega = 1.80$	$\omega = 1.85$	$\omega = 1.90$	$\omega = 1.95$
10^{2}	18	16	5	5	9	24
10^{1}	36	28	31	27	31	67
10^{0}	61	40	65	55	64	94
10^{-1}	62	5'	99	82	75	137
10^{-2}	71	64	133	110	94	191
10^{-3}	85	76	167	138	128	232
10^{-4}	96	88	202	166	144	288

It can be seen from Table 1 that Cudahy's method is initially iteratively faster than either of the other two scans. As the iteration limit becomes finer, however, Cudahy's method becomes iteratively slower. Thus it is likely that the asymptotic rate of convergence of Cudahy's method is less than that for a consistent, cyclic stationary iteration. This assertion cannot be proven experimentally, of course, since the asymptotic rate of convergence is realized only as the number of iterations approaches infinity.

Since the standard scan and the even-odd scan are consistent, the optimum overrelaxation factor can be determined precisely from equation (42). The optimum ω was used to obtain the data shown in Table 1 for these scans. It is interesting to note that the even-odd scan is iteratively faster than the standard scan for all iteration limits shown. Since both scans have the same asymptotic rate of convergence, the even-odd scan can be expected to be iteratively faster than the standard scan for all iteration limits, no matter how fine.

Figs. 6 and 7 show a segment of the data from which Table 1 was constructed, for the standard and even-odd scans. The difference magnitude γ, defined as

$$\gamma \equiv \max \left| u_{i,j}^{(m+1)} - u_{i,j}^{(m)} \right| , \qquad 1 \leq i, j \leq \quad \cdot 1 \tag{58}$$

where N is the number of mesh spacings, is plotted versus the iteration number. For a sufficiently large number of iterations the even-odd scan is iteratively faster than the standard scan by an amount which is represented by a fixed number of iterations at a convergence rate approximately equal to the asymptotic rate of convergence. Since the difference between the scans is a fixed number of iterations, the per

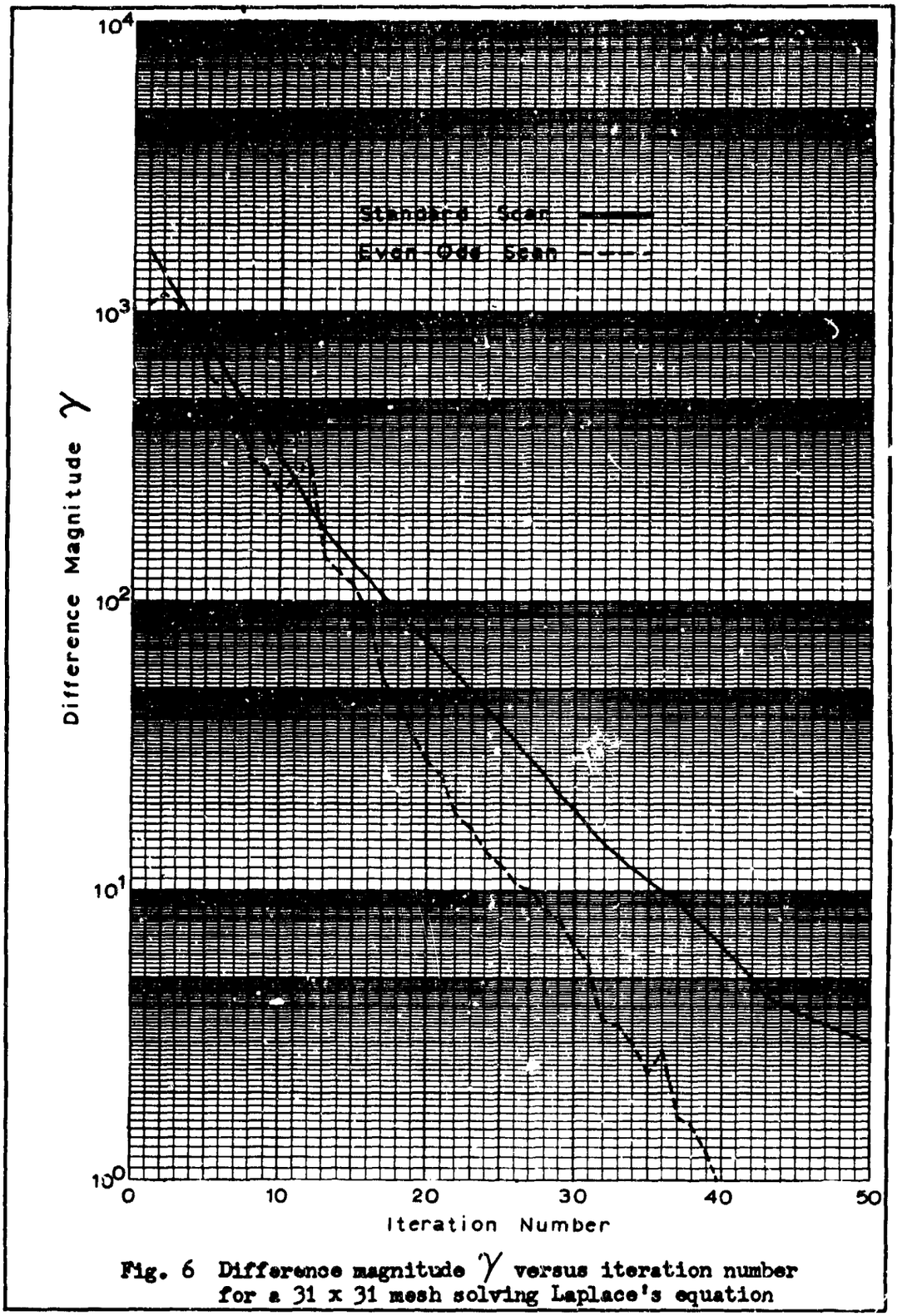

Fig. 6 Difference magnitude γ versus iteration number for a 31 x 31 mesh solving Laplace's equation

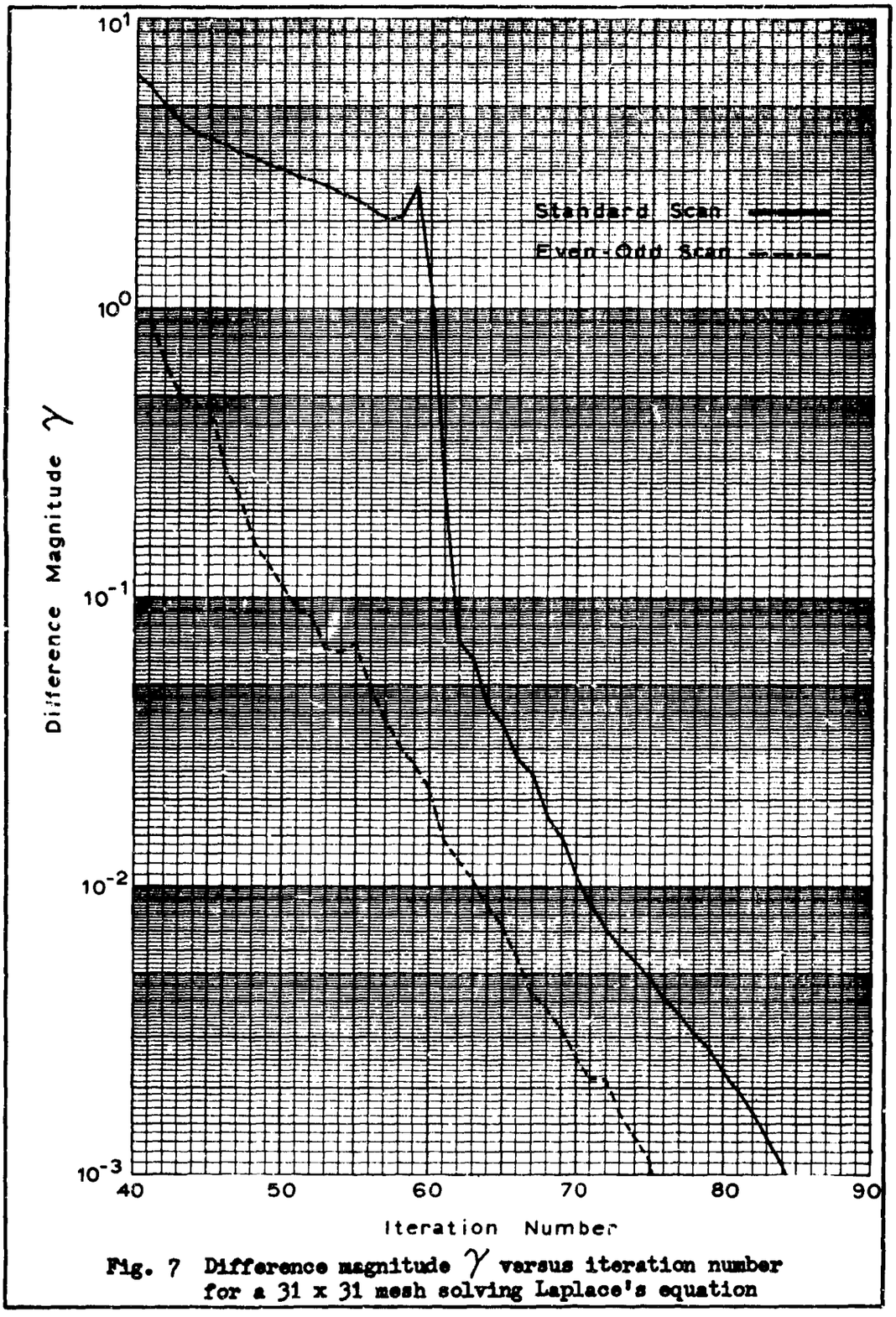

Fig. 7 Difference magnitude γ versus iteration number for a 31 x 31 mesh solving Laplace's equation

cent difference is a function of the number of iterations and will approach zero as the number of iterations approaches infinity.

Fig. 8 shows a comparison between the even-odd scan at optimum ω and Cudahy's method at $\omega = 1.85$. It can be seen that Cudahy's method rapidly approaches its asymptotic rate of convergence, which is not as great as that for the even-odd scan. The somewhat erratic behavior of the difference magnitude γ for the first few iterations illustrates the difficulty in making theoretical predictions of average rates of convergence for a relatively small number of iterations.

To determine whether the mesh size has an important effect on the relative speed between the standard and even-odd scans, the mesh size was varied from a 10 by 10 to a 70 by 70. Computer time limitations precluded inclusion of larger mesh sizes. The results of that comparison are shown in Table 2. Again the optimum ω was calculated from equation (42) and used in these runs. The iteration limit used represents a reduction in the length of the error vector by a factor of approximately 10^{-11}, or eleven orders of magnitude.

The data in Table 2 show that the number of iterations difference between the two scans increases as the mesh spacing decreases, and that the per cent difference, for the same iteration limit, remains approximately the same. This indicates that the relative speed difference is independent of mesh spacing.

Figs. 9 and 10 show the difference magnitude γ as a function of the iteration number for a 40 by 40 mesh. It can be seen that the same general features are present here as were present for the 31 by 31 mesh in Figs. 6 and 7. The only significant change is the difference in asymptotic rate of convergence, the slope of the curve for large iteration numbers.

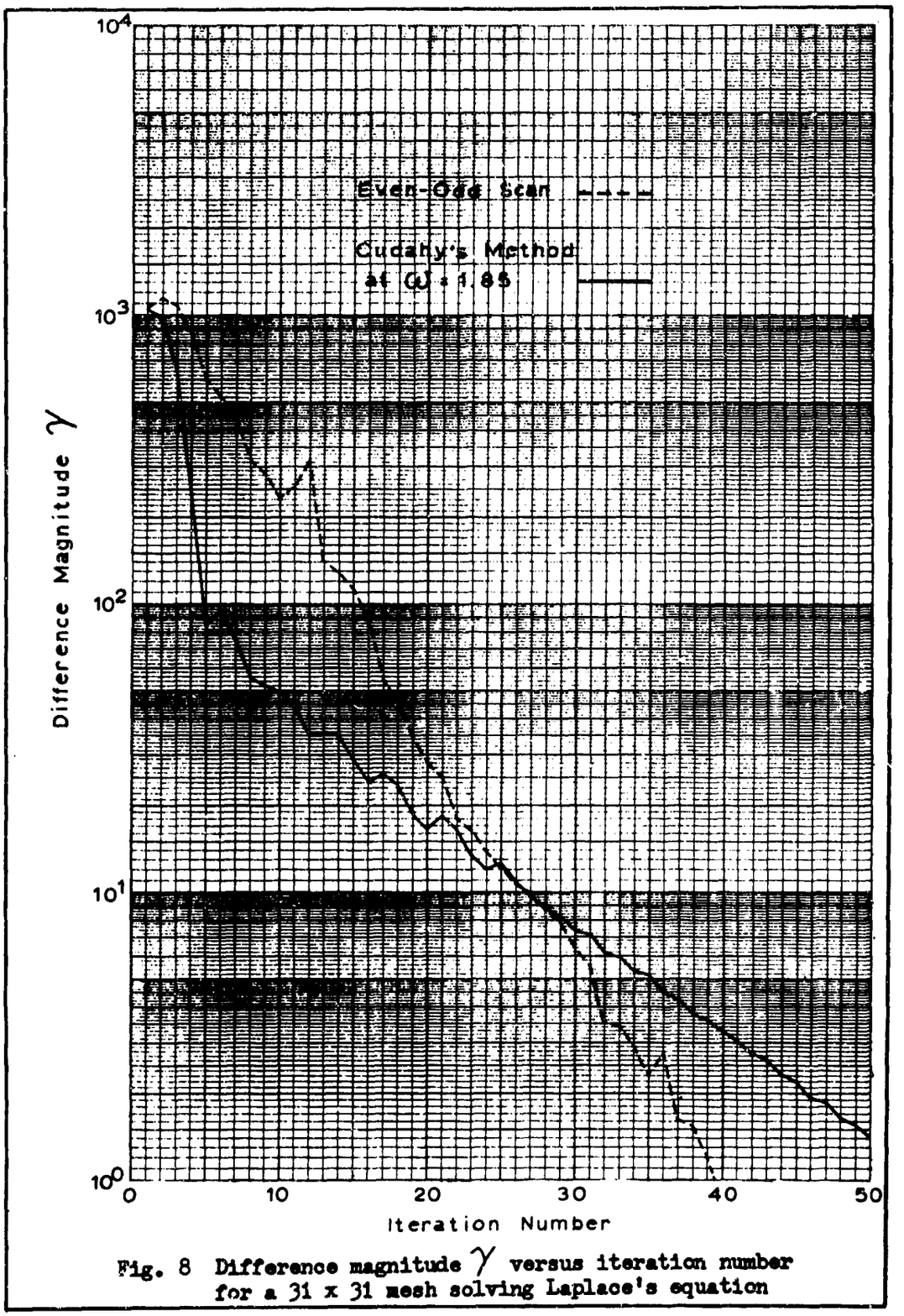

Fig. 8 Difference magnitude γ versus iteration number for a 31 x 31 mesh solving Laplace's equation

Table 2

Number of Iterations Required at an Iteration Limit of 10^{-8} for Solution of Laplace's Equation with Various Mesh Sizes for a Unit Square

Mesh Size	Iterations		Number Difference	Per Cent Difference
	Standard Scan	Even-Odd Scan		
10 x 10	48	45	3	6.3%
20 x 20	98	93	5	5.1%
30 x 30	148	139	9	6.1%
40 x 40	195	186	9	4.6%
50 x 50	244	231	13	5.3%
60 x 60	293	277	16	5.4%
70 x 70	343	323	20	5.8%

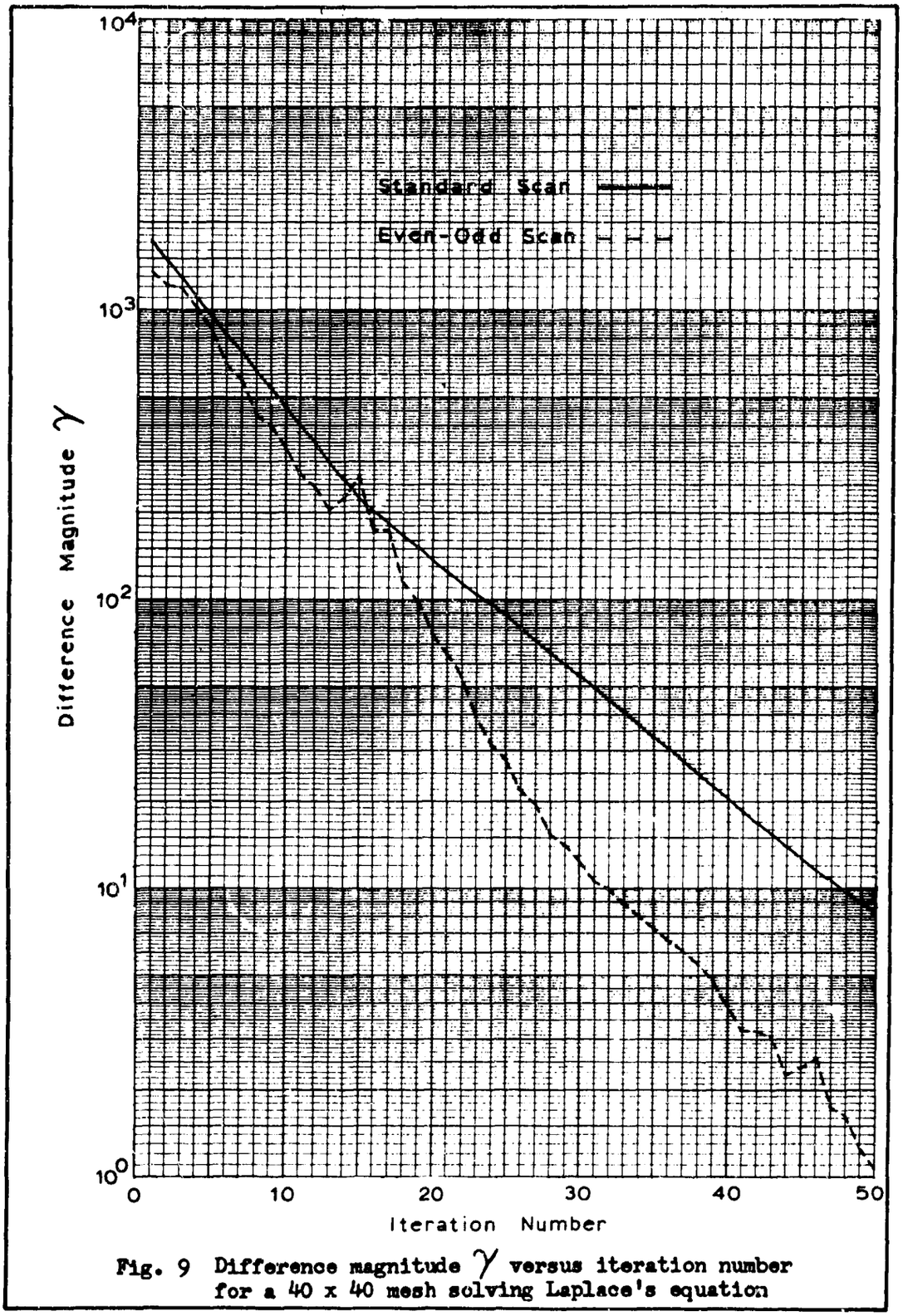

Fig. 9 Difference magnitude γ versus iteration number for a 40 x 40 mesh solving Laplace's equation

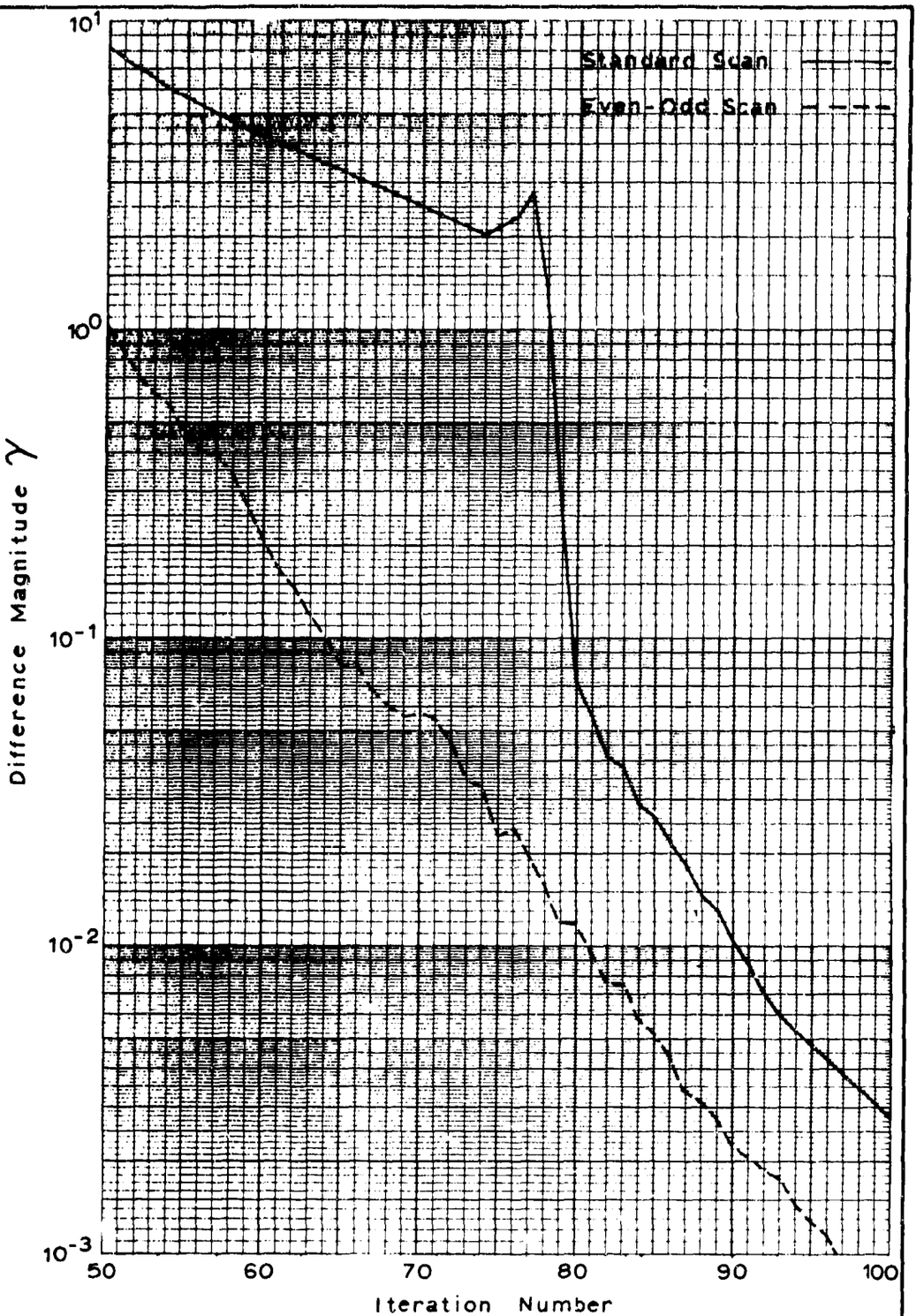

Fig. 10 Difference magnitude γ versus iteration number for a 40 x 40 mesh solving Laplace's equation

Heat Equation

The point equation for the finite difference approximation to the heat equation, corresponding to the matrix equation (9), is given by

$$U(i,j,t+\Delta t) = (1-\omega)U(i,j,t+\Delta t) + \omega \left(\frac{r}{1+4r} \left[U(i+1,j,t+\Delta t) + U(i-1,j,t+\Delta t) + U(i,j+1,t+\Delta t) + U(i,j-1,t+\Delta t) \right] + \frac{1}{1+4r} \left[U(i,j,t) + \beta \Delta t \right] \right) \quad (59)$$

where the (m + 1)'s and (m)'s have been suppressed to allow this equation to represent any ordering.

Solutions of equation (59) by point successive overrelaxation were obtained using the standard and even-odd scans and Cudahy's method. The problem selected was similar to that for Laplace's equation. The boundary values were set at U = 1000, the thermal diffusivity α was taken as $\alpha = 1$, and there was no heat generation ($\beta = 0$). This corresponds to the problem for which Cudahy reported the greatest improvement in rate of convergence for his method compared to the standard scan. The initial values inside the mesh were U = 0 for each mesh point. A uniformly spaced 51 by 51 mesh was employed.

Tables 3, 4, and 5 show the number of iterations required for various iteration limits for $\Delta t = 0.01$, 0.1, and 0.5 respectively. These results correspond closely with those given by Cudahy (Ref 2) for his method and the standard scan up to an iteration limit of 10^{-1}. The present data have been extended to much finer iteration limits, however, in order to indicate the asymptotic behavior of the methods. The even-odd scan, which was found to be iteratively faster than the

Table 3

Number of Iterations Required for Various Iteration Limits for a 51 x 51 Mesh Solving the Heat Equation with $\Delta t = 0.01$

Iteration Limit	Iterations		
	Standard Scan	Even-Odd Scan	Cudahy's Method
	$\omega = 1.25$	$\omega = 1.25$	$\omega = 1.35$
10^{2}	3	3	3
10^{1}	6	5	4
10^{0}	9	7	6
10^{-1}	12	8	8
10^{-2}	15	10	10
10^{-3}	18	11	12
10^{-4}	21	13	14
10^{-5}	25	15	16
10^{-6}	28	16	18
10^{-7}	31	18	21
10^{-8}	35	20	23
10^{-9}	38	21	25
10^{-10}	41	23	28
10^{-11}	44	25	31
10^{-12}	48	26	33

Table 4

Number of Iterations Required for Various Iteration Limits for a 51 x 51 Mesh Solving the Heat Equation with $\Delta t = 0.1$

Iteration Limit	Iterations				
	Standard Scan	Even-Odd Scan	Cudahy's Method		
	ω=1.64	ω=1.64	ω=1.68	ω=1.70	ω=1.72
10^{2}	9	7	4	4	4
10^{1}	18	12	7	7	7
10^{0}	28	17	12	11	11
10^{-1}	38	22	18	17	16
10^{-2}	48	27	25	23	20
10^{-3}	58	32	32	29	27
10^{-4}	68	37	39	35	32
10^{-5}	78	42	47	42	41
10^{-6}	89	47	55	50	53
10^{-7}	100	53	64	60	66
10^{-8}	102	59	73	70	79
10^{-9}	105	63	81	82	95
10^{-10}	109	68	90	96	108
10^{-11}	112	73	99	108	115
10^{-12}	116	78	111	115	123

Table 5

Number of Iterations Required for Various Iteration Limits for a 51 x 51 Mesh Solving the Heat Equation with $\Delta t = 0.5$

Iteration Limit	Iterations		
	Standard Scan	Even-Odd Scan	Cudahy's Method
	$\omega = 1.79$	$\omega = 1.79$	$\omega = 1.85$
10^{2}	16	12	5
10^{1}	34	22	11
10^{0}	52	31	23
10^{-1}	69	43	41
10^{-2}	90	55	61
10^{-3}	101	67	90
10^{-4}	104	79	108
10^{-5}	112	91	117
10^{-6}	123	103	131
10^{-7}	135	115	144
10^{-8}	148	127	160
10^{-9}	160	140	178

standard scan for Laplace's equation, is seen here to be iteratively faster for the heat equation as well.

Again, as was the case for Laplace's equation, Cudahy's method is initially iteratively faster than either of the two consistent, cyclic, stationary scans. The introduction of the factor Δt in the heat equation has the effect of expanding the number of iterations, in relation to a fixed convergence criteria, for which Cudahy's method is faster than the standard scan. Comparing the plots for the standard scan in Figs. 7 and 10 with the data presented in Table 4, it can be seen that the sharp drop in the difference magnitude γ, which for Laplace's equation occured at an iteration limit of approximately 10^0, occurs for a $\Delta t = 0.1$ at an iteration limit of approximately 10^{-7}. Cudahy's method is faster than the standard scan so long as comparisons are confined to this region, and not extended to finer iteration limits. The even-odd scan, on the other hand, is not adversely affected by the transition to a small Δt.

IV. Discussion and Conclusions

Consistent Scans

As shown by the tables and figures in chapter III, the even-odd scan was found to be iteratively faster than the standard scan for all problems tested. This effect can be explained in terms of the theory presented in chapter II by a consideration of the expansion of the error vector at the (m)th iteration in terms of the eigenvalues and eigenvectors of the iteration matrix.

For a sufficiently large number of iterations the error vector following the (m)th iteration is given by equation (20) as

$$\bar{\epsilon}^{(m)} = a_1 \lambda_1^m \bar{e}_1 \tag{20}$$

This condition can be found to be satisfied experimentally when the ratio of the difference magnitude γ for the (m)th iteration to the difference magnitude γ for the (m - 1)th iteration is approximately equal to the spectral radius and when this value is essentially constant for several preceding and several following iterations. For the problems tested this condition was found to apply for all iterations past the 100th iteration. Expressed another way, λ_2^{100} was found to be much less than λ_1^{100}.

In examining the terms in equation (20) it will be recalled that the dominant eigenvector $\bar{e}_1$ was assumed to be normalized. The largest eigenvalue λ_1 is the same, and minimum, for all consistent orderings. This leaves only the quantity a_1, which is the magnitude of the projection of the initial error vector $\bar{\epsilon}^{(0)}$ onto the dominant eigenvector, to account for the observed behavior.

From an examination of equation (20) one would predict that if the quantity a_1 were in fact different for two consistent orderings then the error vector after a sufficient number of iterations would be greater or less for the one scan as compared to the other by a fraction equal to the ratio of the two a_1's. Furthermore, one could predict that the magnitude of the error vector for the iteratively slower scan after $m + m'$ iterations would be equal to that for the faster scan after m iterations, and that m' would be independent of m. Referring to Fig. 10, this is exactly the behavior observed, from which it is concluded that the quantity a_1, corresponding to the even-odd scan is less than the a_1 corresponding to the standard scan. This observation holds for all the problems tested.

That the a_1's could be different for different consistent scans is not surprising since, although all of the eigenvalues are the same, the eigenvectors are known to be, in general, different for different scans (Ref 9:403), and hence the projection of the initial error vector onto the dominant eigenvector will in general be different.

The data indicate that the even-odd scan is preferable, in general, over the standard scan for point successive overrelaxation. The time savings achieved will vary with the size of the problem and/or the iteration limit required, and will be reduced asymptotically to zero as the number of iterations increases toward infinity. The maximum savings observed for few iterations was on the order of 15 to 20%. This value is not unreasonable for the savings to be expected for a small problem (40 x 40 mesh) with a required error vector reduction of three or four orders of magnitude.

Cudahy's Method

One of the initial objectives of this study was to explain why Cudahy's method was faster than the standard scan and then to apply it, or a variation of the principal with block overrelaxation, to three dimensions. Unfortunately, it was discovered that Cudahy's method was not in fact asymptotically faster and hence the expansion to three dimensions was not attempted.

From the data presented in chapter III, Cudahy's method can be seen to be best suited for the transient heat equation at small Δt. The iterative speed of his method can be seen from Tables 3, 4, and 5 to be quite favorable when compared to the standard scan and less so when compared to the even-odd scan. The data in Table 4, for $\Delta t = 0.1$, include the results for Cudahy's method with three different relaxation factors. It can be seen for an iteration limit of 10^{-1} that the highest relaxation factor gives the greatest iterative speed. For finer convergence criteria, however, the ω corresponding to the best speed decreases; and hence, the iterative speed is a function of the iteration limit as well as the relaxation factor.

The method developed by Cudahy was designed to accelerate the convergence of point successive overrelaxation by scanning the boundary values into the mesh as quickly as possible. Each of the four scans employed takes the values from one of the sides of the mesh and distributes them throughout the entire mesh. It can be seen from Fig. 8 that this produces an initial rapid convergence, as seems reasonable, but that the method is eventually iteratively slower than the consistent scans, due to its slower asymptotic rate of convergence.

Block Overrelaxation

Although this present work was confined to point successive overrelaxation, it is known that the method of block successive overrelaxation is asymptotically faster by approximately 40 per cent (Ref 6:205).

Much of the theory presented in chapter II for point successive overrelaxation is also applicable to block overrelaxation. In particular the definition of and techniques for finding consistent orderings are identical. It is only necessary to partition the matrix $\underline{A}$ of equation (1) in the form

$$\underline{A} = \begin{bmatrix} \underline{A}_{1,1} & \underline{A}_{1,2} & \cdots & \underline{A}_{1,n} \\ \underline{A}_{2,1} & \underline{A}_{2,2} & & \underline{A}_{2,n} \\ \vdots & & & \\ \underline{A}_{n,1} & \underline{A}_{n,2} & \cdots & \underline{A}_{n,n} \end{bmatrix} \tag{60}$$

where the diagonal submatrices $\underline{A}_{i,i}$, $1 \leq i \leq N$, are square and nonsingular. The associated matrix $\underline{B}$ is then the block Jacobi matrix $\underline{B}$ defined by $\underline{B} \equiv -\underline{D}^{-1}\ \underline{A} + \underline{I}$.

As a practical matter, the diagonal submatrices of $\underline{A}$ of equation (60) must be tridiagonal so that they may be solved directly.

The relevance of the present work to block successive overrelaxation is that it indicates that it may be possible to find a consistent ordering which is iteratively faster than the standard method of line overrelaxation. The theory does not specifically exclude this possibility; however, the requirement that the submatrices be tridiagonal may make the search difficult.

Recommendations

Since the asymptotic rate of convergence for point and block successive overrelaxation has been shown by Varga to be optimum for consistent orderings (Ref 6:125), and since this study has indicated that the iterative speed difference between two consistent orderings can be significant for small problems, the author feels that further research into accelerating the rate of convergence could be profitably directed toward finding that consistent ordering which yields the greatest initial average rate of convergence.

Extension of the method to three dimensions and testing of various consistent scans in that domain should prove worth-while.

Other iterative methods, such as the alternating-direction implicit methods and semi-iterative methods, have been developed which appear to be faster than point successive overrelaxation, and in some cases, block successive overrelaxation. In particular, Varga has reported that for the model problem (Laplace's equation in a rectangle) the alternating-direction implicit methods have produced convergence rates as much as 38 times as great as for point successive overrelaxation (Ref 6:228). Varga shows that these methods derive their speed primarily from the use of different accelerating parameters on each iteration (Ref 6:217). A significant disadvantage of these methods however, in addition to their complexity, is the fact that they are not rigorously applicable to irregular geometries.

An attempt was made during the present study to increase the rate of convergence for the even-odd scan by the use of non-stationary iterations by assigning a slightly larger than optimum relaxation factor to the even points of the scan and a slightly smaller than

optimum factor for the odd points and by decreasing the difference for subsequent iterations, but the method proved to be initially divergent and was not investigated further. It may be possible to develop a more sophisticated method which would succeed.

Use of the techniques employed by Cudahy, combined with the use of different relaxation factors on each iteration might lead to greater rates of convergence, but this procedure is most likely to succeed as a variation of the alternating-direction implicit methods. Since Cudahy's method has the ability to quickly eliminate large gradients in the approximation to Laplace's equation, it might favorably be combined in an iterative procedure with the even-odd scan, wherein Cudahy's method is applied initially to smooth out large gradients, and then the even-odd scan, with its greater asymptotic rate of convergence, is used to obtain the desired accuracy throughout the mesh.

Bibliography

1. Clark, M. and Hansen, K. F. Numerical Methods of Reactor Analysis. New York: Academic Press, 1964.

2. Cudahy, G. F. "Investigation of Accelerating the Convergence of an Implicit Numerical Solution of Transient Heat Transfer Problems." Masters Thesis, Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, August 1965.

3. Forsythe, G. E. and Wasow, W. R. Finite Difference Methods for Partial Differential Equations. New York: John Wiley and Sons, Inc., 1960.

4. Frankel, S. P. "Convergence rates of Iterative Treatments of Partial Differential Equations." Math. Tables Aids Comput, 4: 65-75 (1950).

5. Smith, G. D. Numerical Solution of Partial Differential Equations. London: Oxford University Press, 1965.

6. Varga, R. S. Matrix Iterative Analysis. New Jersey: Prentice-Hall, Inc., 1962.

7. Young, D. M. "Iterative Methods for Solving Partial Difference Equations of Elliptic Type." Trans. Amer. Math. Soc., 76, 92-111 (1954).

8. Young, D. M. "ORDVAC Solutions of the Dirichlet Problem." Journal of the Association for Computing Machinery, 2: 137-161 (1955).

9. Young, D. M. "The Numerical Solution of Elliptic and Parabolic Partial Differential Equations." Survey of Numerical Analysis, Edited by J. Todd. New York: McGraw-Hill Book Company, Inc. 1962.

10. Wright, K. D. "Investigation of Block Overrelaxation for Accelerating the Convergence of the Implicit Numerical Solution of the Transient Heat Transfer Equation." Masters Thesis, Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, June 1967.

GSP/PH/68-5

Vita

Richard Ernest Durrett was born on 15 June 1940 in Fort Worth, Texas. He gr ated from high school in Meriam, Kansas in 19 and attended the University of Kansas from which he received the gree of Bachelor of Science in Aerospace Engineering and a commission in the USAF in 1963. After serving as a special assistant to the Vice Commander, Air Force Logistics Command, for a period of one year, he became a project engineer in the Service Engine ring Divi ion, Headquarters Oklahoma City Air Mat ial Area, Tinker Air Force Base, Oklahoma served in this position until June, 1966, and then entered the graduate program in Space Physics at the Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio.

Permanent addr 31 West 76th Street
Fairie Village, Kansas

This thesis was typed by Mrs. Carolyn G. Durrett

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Air Force Institute of Technology (AFIT-SE) Wright-Patterson AFB, Ohio 45433	Unclassified
	2b. GROUP

3. REPORT TITLE

The Effect of Ordering on the Average Rate of Convergence of the Point Successive Overrelaxation Iterative Method

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

AFIT Thesis

5. AUTHOR(S) *(First name, middle initial, last name)*

Richard E. Durrett
Capt USAF

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
February 19	56	10

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. N/A	GSP/PH/68-5
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. DISTRIBUTION STATEMENT This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of the Dean of Engineering, Air Force Institute of Technology (AFIT-SE) Wright-Patterson Air Force Base, Ohio 45433

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

The point successive overrelaxation iterative method is applied to the five point finite difference approximations to Laplace's equation and the heat equation using two consistent orderings and one non-stationary iterative procedure. The consistent orderings are found to be asymptotically faster than the non-stationary procedure and to display different iterative speeds. The ordering denoted odd-even parity is found to be 20% faster than the natural ordering for an error vector reduction of three orders of magnitude. An increase in the orders of magnitude reduction produces a proportionate decrease in the per cent difference between the two consistent orderings.

DD FORM 1 NOV 68 1473

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Successive Overrelaxation						
Consistent Ordering						
Rate of Convergence						
Iterative Procedure						